

FMRI Analysis

Experiment Design



Scanning



Pre-Processing



Individual Subject Analysis

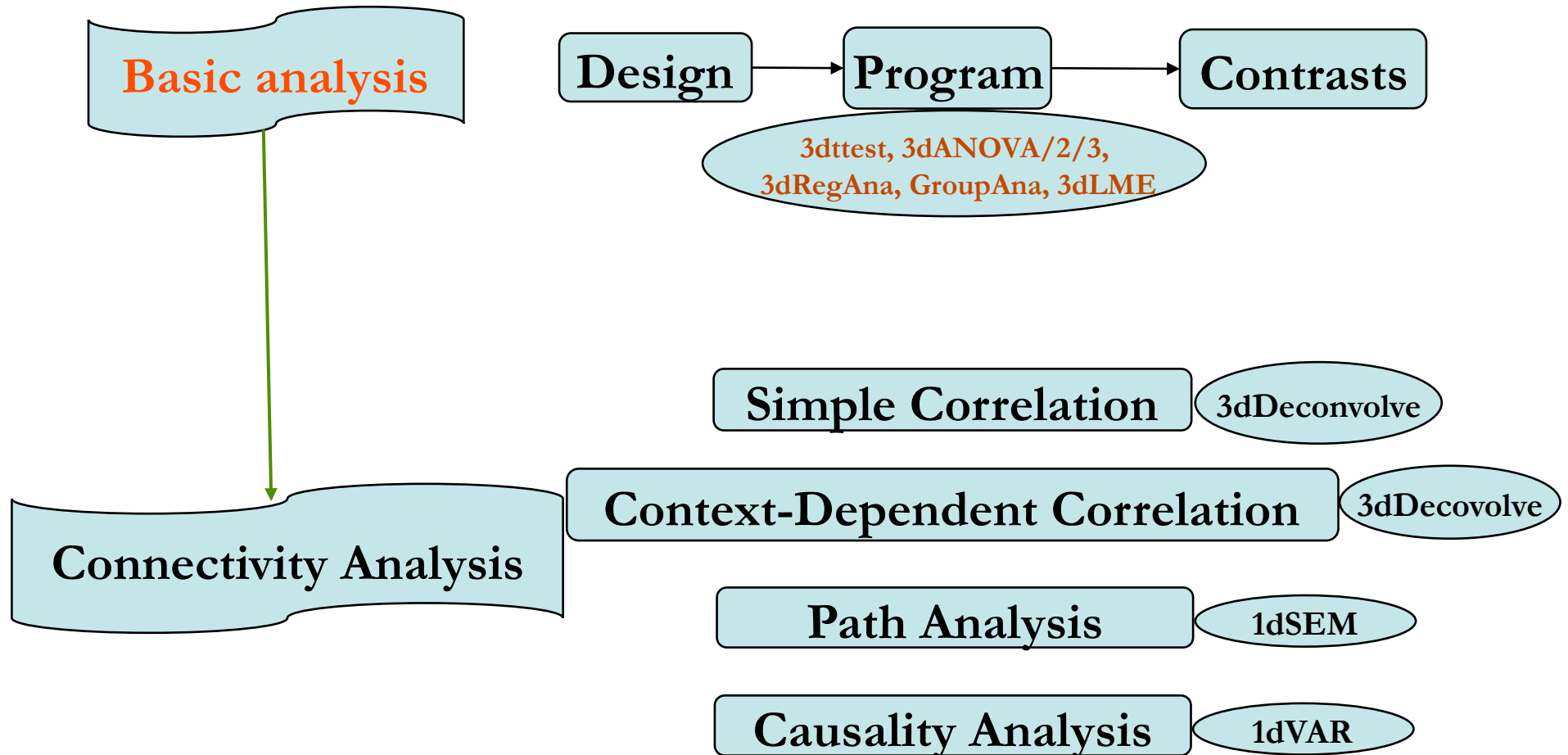


Group Analysis



Post-Processing

Group Analysis



- Group Analysis: Why and how?

- 👉 Group analysis

- ⌞ Make general conclusions about some population, *e.g.*,
 - Do men and women differ on responding to fear?
 - What regions are related to happiness, sad, love, faith, empathy, etc.?
 - ⌞ Partition/untangle data variability into various effects, *e.g.*,
 - What differs when a person listens to classical music vs. rock 'n' roll?

- 👉 Why two tiers of analysis: individual and then group?

- ⌞ No perfect approach to combining both into a batch analysis
 - ⌞ Each subject may have slightly different design or missing data
 - ⌞ High computation cost
 - ⌞ Usually we take β 's (% signal change) to group analysis
 - Within-subject variation relatively small compared to cross-subject

- Group Analysis: Basic concepts

- ☞ Variables

- ↯ Dependent: percent signal changes (β 's)

- ↯ Independent: factors (condition/task, sex, subject) and covariates (IQ, age)

- ☞ Factor: a categorization (variable) of conditions/tasks/subjects

- ↯ 2 types: fixed and random

- ☞ Fixed factor

- ↯ Treated as a fixed variable in the model

- Categorization of experiment conditions (mode: Face/House)

- Group of subjects (male/female, normal/patient)

- ↯ All levels of the factor are of interest and included for replications among subjects

- ↯ Fixed in the sense of inference

- apply only to the specific levels of the factor, e.g., the response to face/house is well-defined

- don't extend to other potential levels that might have been included, e.g., the response to face/house doesn't say anything about the response to music

- Group Analysis: Basic concepts

- ☞ Random factor

- ↯ Exclusively refers to subject in FMRI
 - ↯ Treated as a random variable in the model
 - average + random effects uniquely attributable to each subject: $N(\theta, \sigma^2)$
 - ↯ Each subject is of NO interest
 - ↯ Random in the sense of inference
 - subjects serve as a random sample of a population
 - this is why we recruit a lot of subjects for a study
 - inferences can be generalized to a population
 - we usually have to set a long list of criteria when recruiting subjects (right-handed, healthy, age 20-40, native English speaker, etc.)

- ☞ Covariates

- ↯ Confounding/nuisance effects
 - Continuous variables of no interest
 - May cause spurious effects or decrease power if not modeled
 - Some measures about subject: age, IQ, cross-conditions/tasks behavior data, etc.

- Group Analysis: Types

- ☞ Fixed: factor, analysis/model/effects

- Fixed-effects analysis (sometimes): averaging among a few subjects

- ☞ Non-parametric tests

- ☞ Mixed design

- Mixed design: crossed [e.g., AXBXC] and nested [e.g., BXC(A)]

- Psychologists: Within-subject (repeated measures) / between-subjects factor

- ☞ Mixed-effects analysis (aka random-effects)

- ↙ ANOVA: contains both types of factors: both inter/intra-subject variances

- Crossed, e.g., AXBXC

- Nested, e.g., BXC(A)

- ↙ ANCOVA

- ↙ LME

- Unifying and extending ANOVA and ANCOVA

- Using ML or ReML

• Group Analysis: What do we get out of the analysis

☞ Using an intuitive example of income (dependent variable)

☞ Factor A: sex (men vs. women)

☞ factor B: race (whites vs. blacks)

☞ Main effect

☞ F : general information about all levels of a factor

☞ Any difference between two sexes or races

➤ men > women; whites > blacks

☞ Is it fair to only focus on main effects?

☞ Interaction

➤ F : Mutual/reciprocal influence among 2 or more factors

➤ Effect of a factor depends on levels of other factors, e.g.,

➤ Black men < black women

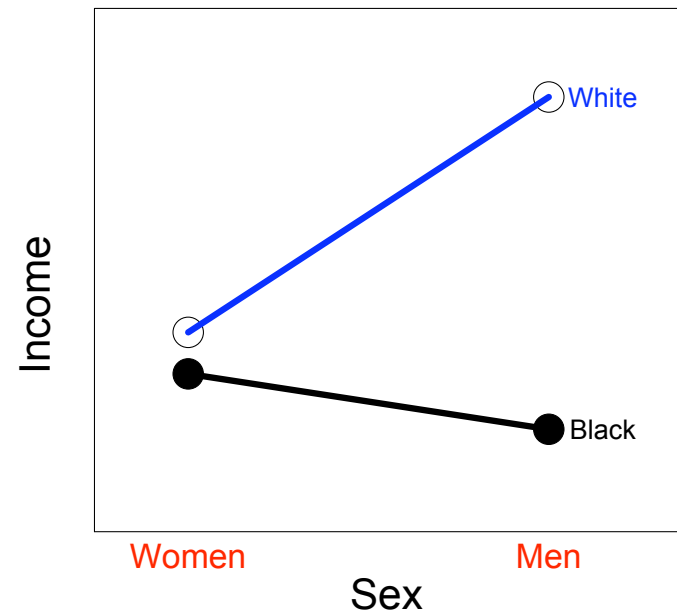
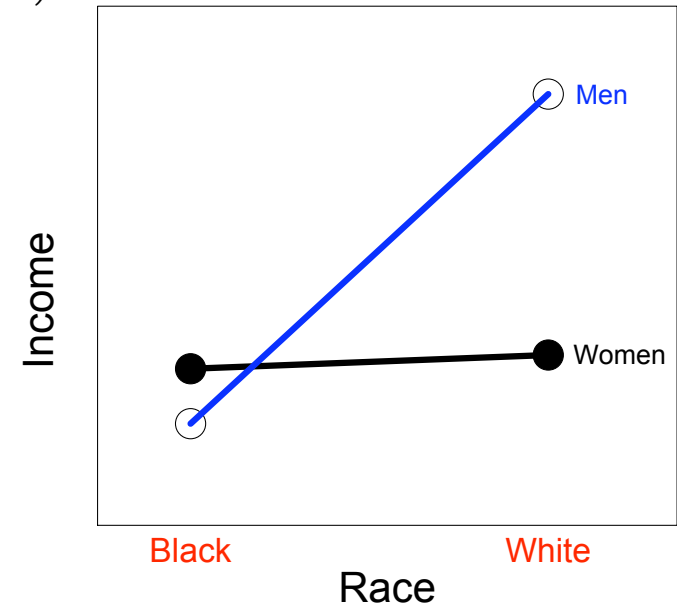
➤ Black women almost the same as white women

➤ Black men << white men

☞ General linear test

➤ Contrast

➤ General linear test (e.g., trend analysis)



- Group Analysis: Types

- ☞ Averaging across subjects (fixed-effects analysis)

- ↯ Number of subjects $n < 6$
 - ↯ Case study: can't generalize to whole population
 - ↯ Simple approach (**3dcalc**)
 - $T = \sum t_{ii} / \sqrt{n}$
 - ↯ Sophisticated approach
 - $B = \sum(b_i / \sqrt{v_i}) / \sum(1 / \sqrt{v_i})$, $T = B \sum(1 / \sqrt{v_i}) / \sqrt{n}$, v_i = variance for i -th regressor
 - $B = \sum(b_i / v_i) / \sum(1 / v_i)$, $T = B \sqrt{[\sum(1 / v_i)]}$
 - Combine individual data and then run regression

- ☞ Mixed-effects analysis

- ↯ Number of subjects $n > 10$
 - ↯ Random effects of subjects
 - ↯ Individual and group analyses: separate
 - ↯ Within-subject variation ignored
 - ↯ Main focus of this talk

- Group Analysis: Programs in AFNI

- ☞ Non-parametric analysis

- ↯ $4 < \text{number of subjects} < 10$
 - ↯ No assumption of normality; statistics based on ranking
 - ↯ Programs
 - **3dWilcoxon** (\sim paired t -test)
 - **3dMannWhitney** (\sim two-sample t -test)
 - **3dKruskalWallis** (\sim between-subjects with **3dANOVA**)
 - **3dFriedman** (\sim one-way within-subject with **3dANOVA2**)
 - **Permutation test**
 - ↯ Multiple testing correction with FDR (**3dFDR**)
 - ↯ Less sensitive to outliers (more robust)
 - ↯ Less flexible than parametric tests
 - ↯ Can't handle complicated designs with more than one fixed factor

- **Group Analysis: Programs in AFNI**

- ☞ Parametric tests (mixed-effects analysis)

- ↙ Number of subjects > 10

- ↙ Assumption: Gaussian random effects

- ↙ Programs

- **3dttest** (one-sample, two-sample and paired t)

- **3dANOVA** (one-way between-subject)

- **3dANOVA2** (one-way within-subject, 2-way between-subjects)

- **3dANOVA3** (2-way within-subject and mixed, 3-way between-subjects)

- **3dRegAna** (regression/correlation, simple unbalanced ANOVA, simple ANCOVA)

- **GroupAna** (Matlab package for up to 5-way ANOVA)

- **3dLME** (R package for all sorts of group analysis)

- Group Analysis: Planning for mixed-effects analysis

- ☞ How many subjects?

- ↙ Power/efficiency: proportional to \sqrt{n} ; $n > 10$
 - ↙ Balance: Equal number of subjects across groups if possible

- ☞ Input files

- ↙ Common brain in tlrc space (resolution doesn't have to be 1x1x1 mm³)
 - ↙ Percent signal change (**not** statistics) or normalized variables
 - HRF magnitude: Regression coefficients
 - Linear combinations of β 's

- ☞ Analysis design

- ↙ Number of factors
 - ↙ Number of levels for each factor
 - ↙ Factor types
 - Fixed (factors of interest) vs. random (subject)
 - Cross/nesting: Balanced? Within-subject/repeated-measures vs. between-subjects
 - ↙ Which program?
 - **3dttest, 3dANOVA/2/3, GroupAna, 3dRegAna, 3dLME**

- **Group Analysis: Planning**

- ✎ **Thresholding**

- ↯ Two-tail by default in AFNI
 - ↯ If one-tail p is desirable, look for $2p$ on AFNI

- ✎ **Scripting – 3dANOVA3**

- ↯ **Three-way between-subjects (type 1)**

- 3 categorizations of groups: sex, disease, age

- ↯ **Two-way within-subject (type 4):** Crossed design $A \times B \times C$

- One group of subjects: 16 subjects
 - Two categorizations of conditions: A – category; B - affect

- ↯ **Two-way mixed (type 5):** $B \times C(A)$

- Nesting (between-subjects) factor (A): subject classification, e.g., sex
 - One category of condition (within-subject factor B): condition (visual vs. auditory)
 - Nesting: balanced

• Group Analysis: Example – 2-way within-subject ANOVA

```
3dANOVA3 -type 4 -alevels 3 -blevels 3 -clevels 16 \
```

Model type,
Factor levels

```
-dset 1 1 1 stats.sb04.beta+tlrc'[0]' \
```

```
-dset 1 2 1 stats.sb04.beta+tlrc'[1]' \
```

```
-dset 1 3 1 stats.sb04.beta+tlrc'[2]' \
```

```
-dset 2 1 1 stats.sb04.beta+tlrc'[4]' \
```

...

```
-fa Category \
```

```
-fb Affect \
```

```
-fab CatXAff \
```

```
-amean 1 T \ (coding with indices)
```

```
-acontr 1 0 -1 TvsF \ (coding with coefficients)
```

```
-bcontr 0.5 0.5 -1 non-neu \ (coefficients)
```

```
-aBcontr 1 -1 0 : 1 TvsE-pos \ (coefficients)
```

```
-Abcontr 2 : 1 -1 0 EPosvsENeg \ (coefficients)
```

```
-bucket anova33
```

Input for each cell in
ANOVA table:
totally 3X3X16 = 144

F tests: Main effects &
interaction

t tests: 1st order Contrasts

t tests: 2nd order
Contrasts

Output: bundled

- Group Analysis: GroupAna

- 👉 Multi-way ANOVA

- ↯ Matlab script package for up to 5-way ANOVA
 - ↯ Can handle both volume and surface data
 - ↯ Can handle up to 4-way unbalanced designs
 - Unbalanced: unequal number of subjects across groups
 - No missing data from subjects allowed
 - ↯ Downsides
 - Requires Matlab plus Statistics Toolbox
 - Slow (minutes to hours): GLM approach - regression through dummy variables
 - Complicated design, and compromised power
 - ↯ Solution to heavy duty computation
 - Input with lower resolution recommended
 - Resample with **adwarp -dxyz #** or **3dresample**
 - ↯ See <http://afni.nimh.nih.gov/sscc/gangc> for more info

- 👉 Alternative: **3dLME**

- Group Analysis: ANCOVA (ANalysis of COVAriances)

- ☞ Why ANCOVA?

- ⌞ Subjects or cross-regressors effects might not be an ideally randomized
 - ⌞ If not controlled, such variability will lead to loss of power and accuracy
 - ⌞ Different from **amplitude modulation**: **cross**-regressors vs. **within**-regressor variation
 - ⌞ Direct control via design: balanced selection of subjects (e.g., age group)
 - ⌞ Indirect (statistical) control: add covariates in the model
 - ⌞ Covariate (variable of no interest): uncontrollable/confounding, usually continuous
 - Age, IQ, cortex thickness
 - Behavioral data, e.g., response time, correct/incorrect rate, symptomatology score, ...

- ☞ ANCOVA = Regression + ANOVA

- ⌞ Assumption: **linear** relation between HDR and the covariate
 - ⌞ GLM approach: accommodate both categorical and quantitative variables

- ☞ Programs

- ⌞ **3dRegAna**: for simple ANCOVA
 - If the analysis can be handled with 3dttest without covariates
 - See <http://afni.nimh.nih.gov/sscc/gangc/ANCOVA.html> for more information
 - ⌞ **3dLME**: R package

• Group Analysis: 3dLME

✎ An R package

- ✎ Open source platform
- ✎ Linear mixed-effects (LME) modeling
- ✎ Versatile: handles almost all situations in one package
 - Unbalanced designs (unequal number of subjects, missing data, etc.)
 - ANOVA and ANCOVA, but unlimited number of factors and covariates
 - Able to handle HRF modeling with basis functions
 - Violation of sphericity: heteroscedasticity, variance-covariance structure
 - Model fine-tuning
- ✎ No scripting (input is bundled into a text file model.txt)
- ✎ Disadvantages
 - High computation cost (lots of repetitive calculation)
 - Sometimes difficult to compare with traditional ANOVA
- ✎ See <http://afni.nimh.nih.gov/sscc/gangc/lme.html> for more information

- Group Analysis: 3dLME

- ✎ HRF modeled with basis functions

- ✎ Traditional approach: AUC

- Can't detect shape difference

- Difficult to handle betas with mixed signs

- ✎ LME approach

- Usually $H_0: \beta_1 = \beta_2 = \dots = \beta_k$

- But now we don't care about the differences among β 's

- Instead we want to detect shape difference

- $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$

- Solution: take all β 's and model with no intercept

- But we have to deal with temporal correlations among β 's!

• Group Analysis: 3dLME

👉 Running LME

🔗 Create a text file `model.txt` (3 fixed factors plus 1 covariate)

```
Data:Volume                                <-- either Volume or Surface
Output:FileName                            <-- any string (no suffix needed)
MASK:Mask+tlrc.BRIK                       <-- mask dataset
Model:Gender*Object*Modality+Age           <-- model formula for fixed effects
COV:Age                                    <-- covariate list
RanEf:TRUE                                <-- random effects
VarStr:0
CorStr:0
SS:sequential
MFace-FFace                               <-- contrast label
Male*Face*0*0-Female*Face*0*0             <-- contrast specification
MVisual-Maudial
Male*0*Visual*0-Male*0*Audial*0
.....
Subj      Gender      Object      Modality      Age      InputFile
Jim        Male        Face        Visual        25      file1+tlrc.BRIK
Carol      Female      House      Audial        23      file2+tlrc.BRIK
Karl       Male        House      Visual        26      file3+tlrc.BRIK
Casey      Female      Face        Audial        24      file4+tlrc.BRIK
.....
```

🔗 Run `3dLME.R MyOut &`

• Group Analysis: 3dLME

☞ Running LME: A more complicated example

↙ HRF modeled with 6 tents

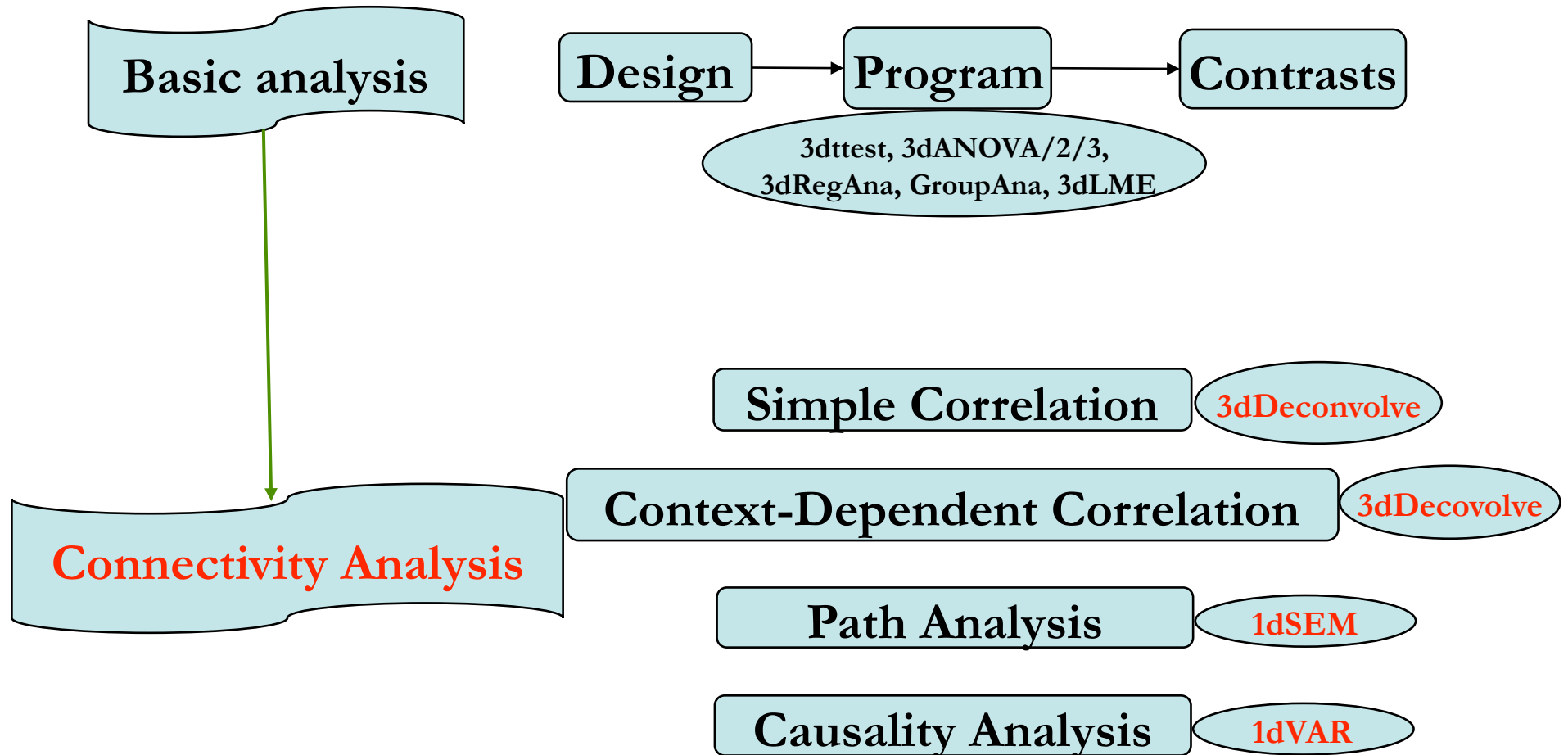
↙ Null hypothesis: no HRF difference between two conditions

```
Data:Volume                <-- either Volume or Surface
Output:test                <-- any string (no suffix needed)
MASK:Mask+tlrc.BRIK       <-- mask dataset
Model:Time-1              <-- model formula for fixed effects
COV:                      <-- covariate list
RanEff:TRUE               <-- random effect specification
VarStr:0                  <-- heteroscedasticity?
CorStr:1~Order|Subj       <-- correlation structure
SS: sequential            <-- sequential or marginal
```

Subj	Time	TimeOrder	InputFile
Jim	t1	1	contrastT1+tlrc.BRIK
Jim	t2	2	contrastT2+tlrc.BRIK
Jim	t3	3	contrast3+tlrc.BRIK
Jim	t4	4	contrast4+tlrc.BRIK
.....			

↙ Output: an F for H_0 , β and t for each basis function

Group Analysis



- Connectivity: Correlation Analysis

- ☞ Correlation analysis (aka functional connectivity)

- ↯ Similarity between a seed region and the rest of the brain
 - ↯ Says not much about causality/directionality
 - ↯ Voxel-wise analysis
 - ↯ Both individual subject and group levels
 - ↯ Two types: **simple** and **context-dependent** correlation (a.k.a. PPI)

- ☞ Steps at individual subject level

- ↯ Create ROI (a sphere around peak t-statistic or an anatomical structure)
 - ↯ Isolate signal for a condition/task
 - ↯ Extract seed time series
 - ↯ Run correlation analysis through regression analysis
 - ↯ More accurately, partial (multiple) correlation

- ☞ Steps at group level

- ↯ Convert correlation coefficients to Z (Fisher transformation): `3dcalc`
 - ↯ One-sample t test on Z scores: `3dttest`

- ☞ More details: <http://afni.nimh.nih.gov/sscc/gangc>

• Connectivity: Path Analysis or SEM

👉 Causal modeling (a.k.a. structural or effective connectivity)

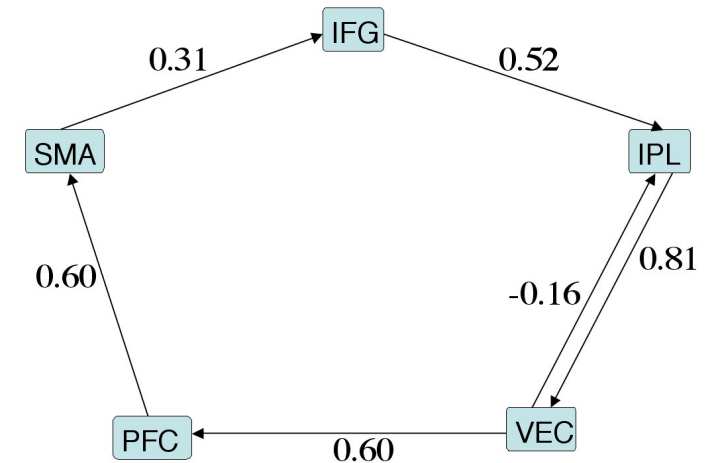
↙ Start with a network of ROI's

↙ Path analysis

- Assess the network based on correlations (covariances) of ROI's
- Minimize discrepancies between correlations based on data and estimated from model
- Input: Model specification, correlation matrix, residual error variances, DF
- Output: Path coefficients, various fit indices

↙ Caveats

- H_0 : It is a good model; Accepting H_0 is usually desirable
- Valid only with the data and model specified
- No proof: modeled through correlation analysis
- Even with the same data, an alternative model might be equally good or better
- If one critical ROI is left out, things may go awry
- Interpretation of path coefficient: NOT correlation coefficient, possible >1



- **Connectivity: Path Analysis or SEM**

- 👉 **Path analysis with 1dSEM**

- ⌞ **Model validation**: ‘confirm’ a theoretical model

- Accept, reject, or modify the model?

- ⌞ **Model search**: look for ‘best’ model

- Start with a minimum model (1): can be empty

- Some paths can be excluded (0), and some optional (2)

- Model grows by adding one extra path a time

- ‘Best’ in terms of various fit criteria

- ⌞ More information <http://afni.nimh.nih.gov/sscc/gangc/PathAna.html>

- 👉 **Difference between causal and correlation analysis**

- ⌞ Predefined network (model-based) vs. network search (data-based)

- ⌞ Modeling: causation (and directionality) vs. correlation

- ⌞ ROI vs. voxel-wise

- ⌞ Input: correlation (condensed) vs. original time series

- ⌞ Group analysis vs. individual + group

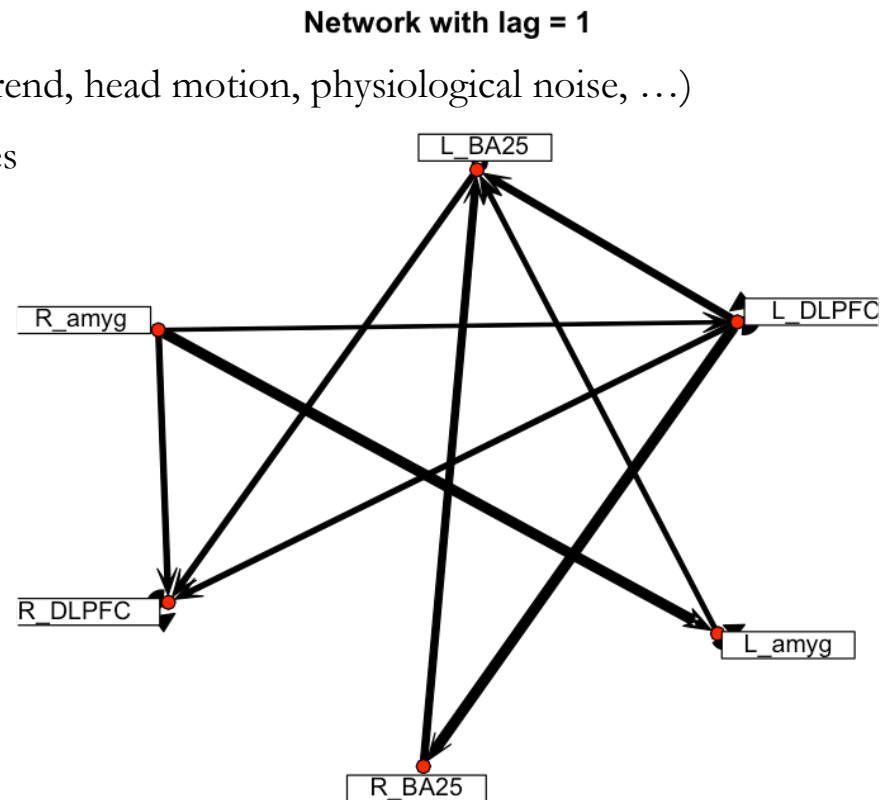
• Connectivity: Causality Analysis or VAR

✎ Causal modeling (a.k.a. structural or effective connectivity)

- ✎ Start with a network of ROI's
- ✎ Causality analysis through vector auto-regressive modeling (VAR)
 - Assess the network based on correlations of ROIs' time series
 - If values of region X provide statistically significant information about future values of Y , X is said to Granger-cause Y
 - Input: time series from ROIs, covariates (trend, head motion, physiological noise, ...)
 - Output: Path coefficients, various fit indices

✎ Causality analysis with **1dVAR**

- ✎ Written in R
- ✎ Run in interactive mode for each subject
- ✎ Generate a network and path matrix
- ✎ A list of model diagnostic tests
- ✎ Run group analysis on path coefficients



- **Connectivity: Causality Analysis or VAR**

- ✎ **Causal modeling** (a.k.a. structural or effective connectivity)

- ⌞ Caveats

- It has assumptions (stationary property, Gaussian residuals, and linearity)
 - Require accurate region selection: missing regions may invalidate the analysis
 - Sensitive to number of lags
 - Time resolution
 - No proof: modeled through statistical analysis
 - Not really cause-effect in strict sense
 - Interpretation of path coefficient: temporal correlation

- ✎ **SEM versus VAR**

- ⌞ Predefined network (model-based) among ROIs
 - ⌞ Modeling: statistical causation (and directionality)
 - ⌞ Input: correlation (condensed) vs. original time series
 - ⌞ Group analysis vs. individual + group

- Connectivity: Causality Analysis or VAR

☞ Why temporal resolution is important?

